

Original Paper

The German Version of the Mobile App Rating Scale (MARS-G): Development and Validation Study

Eva-Maria Messner¹, MA; Yannik Terhorst¹, MSc; Antonia Barke², DPhil; Harald Baumeister¹, PhD, Prof Dr; Stoyan Stoyanov³, MRES; Leanne Hides⁴, PhD, Prof Dr; David Kavanagh⁵, PhD, Prof Dr; Rüdiger Pryss⁶, PhD, Prof Dr; Lasse Sander⁷, PhD; Thomas Probst⁸, PhD, Prof Dr

¹Clinical Psychology and Psychotherapy, Ulm University, Ulm, Germany

²Clinical and Biological Psychology, Catholic University Eichstaett-Ingolstadt, Eichstaett-Ingolstadt, Germany

³Creative Industries Faculty, Queensland University of Technology, Brisbane, Australia

⁴Faculty of Health and Behavioural Sciences, University of Queensland, Brisbane, Australia

⁵Psychology and Counselling, Queensland University of Technology, Brisbane, Australia

⁶Institute of Databases and Information Systems, Ulm University, Ulm, Germany

⁷Department of Rehabilitation Psychology and Psychotherapy, University of Freiburg, Freiburg, Germany

⁸Department for Psychotherapy and Biopsychosocial Health, Danube University Krems, Krems, Austria

Corresponding Author:

Eva-Maria Messner, MA

Clinical Psychology and Psychotherapy

Ulm University

Albert-Einstein-Allee 47

Ulm

Germany

Phone: 49 07315032802

Email: eva-maria.messner@uni-ulm.de

Abstract

Background: The number of mobile health apps (MHAs), which are developed to promote healthy behaviors, prevent disease onset, manage and cure diseases, or assist with rehabilitation measures, has exploded. App store star ratings and descriptions usually provide insufficient or even false information about app quality, although they are popular among end users. A rigorous systematic approach to establish and evaluate the quality of MHAs is urgently needed. The Mobile App Rating Scale (MARS) is an assessment tool that facilitates the objective and systematic evaluation of the quality of MHAs. However, a German MARS is currently not available.

Objective: The aim of this study was to translate and validate a German version of the MARS (MARS-G).

Methods: The original 19-item MARS was forward and backward translated twice, and the MARS-G was created. App description items were extended, and 104 MHAs were rated twice by eight independent bilingual researchers, using the MARS-G and MARS. The internal consistency, validity, and reliability of both scales were assessed. Mokken scale analysis was used to investigate the scalability of the overall scores.

Results: The retranslated scale showed excellent alignment with the original MARS. Additionally, the properties of the MARS-G were comparable to those of the original MARS. The internal consistency was good for all subscales (ie, omega ranged from 0.72 to 0.91). The correlation coefficients (r) between the dimensions of the MARS-G and MARS ranged from 0.93 to 0.98. The scalability of the MARS (H=0.50) and MARS-G (H=0.48) were good.

Conclusions: The MARS-G is a reliable and valid tool for experts and stakeholders to assess the quality of health apps in German-speaking populations. The overall score is a reliable quality indicator. However, further studies are needed to assess the factorial structure of the MARS and MARS-G.

(JMIR Mhealth Uhealth 2020;8(3):e14479) doi: [10.2196/14479](https://doi.org/10.2196/14479)

KEYWORDS

mHealth; Mobile App Rating Scale; mobile app; assessment; rating; scale development

Introduction

Mobile phones are an integral part of modern life. In Europe, 67% of the population owns smartphones, and the number of smartphone users is rising worldwide [1]. It has been reported that 30% of Germans have 11 to 20 apps installed on their smartphones [2]. The use of mobile apps to improve mental health and well-being is becoming increasingly common, with roughly 29% of Germans currently using at least one health app [3].

Globally, between 95 million and 130 million people speak German, making it the 11th most spoken language worldwide [4,5]. Elderly individuals and people with basic education are commonly monolingual in Germany [6]. Yet, these populations have a high need for assistance in developing and maintaining health behaviors and could benefit from the use of mobile health apps (MHAs). A German MHA rating scale could help researchers and health care providers assess the quality of health apps quickly and reliably in their mother tongue. Furthermore, it would be easy to rate a German app with a German scale.

MHAs offer unique and diverse possibilities for health promotion. They allow ecological momentary assessments [7,8] and interventions [8,9]. Additionally, they can be used irrespectively of geographical, financial, and social conditions; can simultaneously target nonclinical and clinical populations; and have the capacity to provide diverse health-management strategies in an ecological setting [10]. Moreover, they support individuals, including those from high-need high-cost populations (eg, those with chronic or lifestyle diseases), in managing their health [9]; reduce help-seeking barriers; and offer a wide range of engagement options [10].

Despite the recent proliferation of MHAs [11], there are no universally accepted criteria for measuring and reporting their quality [8,12]. Therefore, it is necessary to support researchers, users, and health care providers (eg, physicians, psychotherapists, and physiotherapists) in selecting high-quality MHAs. Safe and reliable use of MHAs requires evidence of efficacy and quality, information about data protection, information about routines for emergencies (eg, self-harm and adverse effects), and overall consideration of associated risks [10].

Boudreaux and colleagues [13] suggested the following seven strategies to evaluate MHA quality: (1) Review the scientific literature; (2) Search app clearinghouse websites; (3) Search app stores; (4) Review app descriptions, user ratings, and user reviews; (5) Conduct a social media query within professional and, if available, patient networks; (6) Pilot the apps; and (7) Elicit feedback from users. This process might be too demanding for health care providers and end users when making treatment choices. A standardized and reliable quality assessment tool could facilitate this process.

Several MHA evaluation scales exist to date. The American Psychological Association released an app evaluation model comprising 33 items across the following five scales: background information, risk/privacy and security, evidence, ease of use, and interoperability [12]. The main aim of this

model is to assess the likelihood of harm [9]. However, the validity and reliability assessment of this rating instrument has not yet been reported, and there is no agreement regarding its application [14].

Baumel and colleagues [15] developed the Evaluation Tool for Mobile and Web-Based eHealth Interventions (ENLIGHT), according to a comprehensive systematic review of relevant criteria. The tool allows the evaluation of app quality in terms of seven dimensions (usability, visual design, user engagement, content, therapeutic persuasiveness, therapeutic alliance, and general subjective evaluation) with 28 items. ENLIGHT also provides a checklist to assess credibility, evidence base, privacy explanation, and basic security.

The Mobile App Rating Scale (MARS) [16] is the most commonly used app evaluation tool that allows electronic health experts to rate MHAs. It includes 19 items comprising four subscales on objective MHA characteristics (engagement, functionality, esthetics, and information quality) and a further 10 items comprising two subscales on subjective characteristics (subjective app quality and perceived impact). The subscale and overall scores indicate the quality of MHAs. The MARS has been used to scientifically assess app quality in the following fields: weight management, physical activity, heart failure, diet in children and adolescents, medication adherence, mindfulness, back pain, chronic pain, smoking cessation, and depression [8,17–24]. Thus, it is the most widely used MHA quality rating tool in the scientific community. Furthermore, numerous international efforts promoting safe MHA use (eg, Mobile Health App Database, PsyberGuide or App Script, Reachout, Kinds Helpline, Health Navigator, and Vic Health) are based on the MARS.

The original version of the MARS is in English, but culture- and language-specific app ratings are needed globally. Spanish and Italian versions of the MARS have been developed [25,26]. A German MARS is necessary considering the growing and unregulated MHA market in Germany. Therefore, this study aimed to develop and validate a German version of the Mobile App Rating Scale (MARS-G) and to investigate the scalability of the overall MARS score with Mokken scale analysis—an approach that is closely related to item response theory.

Methods

Adaptation and Translation

The MARS was translated from English into German by two independent bilingual scientists (EMM and TP). After review and discussion of both forward translations, a pilot version of the MARS-G was created. This pilot version underwent blind back translation by two bilingual speakers with different backgrounds (a postdoctoral psychologist [AB] and a nonacademic individual [LMZ]). Thereafter, the back translation was compared with the original English version by the bilingual scientists (EMM and TP), and the penultimate version of the MARS-G was created. This version was evaluated for comprehensibility by three researchers and three nonacademics. After addressing their comments, the final version of the MARS-G was created and used in this study. The MARS-G can

be downloaded from the supplementary materials or obtained from the authors on request.

Search and Procedure

The MARS-G was validated within the framework of a study on the quality of apps targeting anxiety (E M Messner et al, unpublished data, 2020). Apps were identified using the following search terms: anxiety, fears, anxiety attack, anxious, anxiousness, anxiety disorder, fear, dread, fearful, panic, panic attacks, worry, and worries. Each search term was provided separately, as no truncation or use of logic operators (AND, OR, and NOT) was possible in the Google Play Store and iOS Store.

The inclusion process was divided into three steps (searching, screening, and determining eligibility). (1) Using the search terms mentioned above, the initial app pool was identified. (2) App details on the store sites were screened, and apps were downloaded and reviewed if they were developed for anxiety, were available in German or English, were downloadable through the official Google Play Store or iOS Store, and met no relevant exclusion criteria (app bundles [many applications only available as a group]). (3) All downloaded apps were assessed and excluded if they did not address anxiety, were not in German or English, were malfunctioning, or met relevant exclusion criteria (device incompatibility and development/test phase). We identified 3562 MHAs from the app stores. However, we excluded 810 duplicate apps, 2577 apps considered inappropriate on screening, and 71 apps considered ineligible. The remaining 104 apps were rated using the MARS and MARS-G by two independent trained raters. The raters tested all MHAs for 15 minutes. Quality was assessed immediately after the testing period in both languages. The assessment of the MARS-G is present in a review evaluating the quality of MHAs available for anxiety (E M Messner et al, unpublished data, 2020).

Rater Training

We followed the rating methodology in the original study by Stoyanov and colleagues [16]. We created a YouTube video with an introduction on MARS-G rating and an exercise on how to rate an app used as an exemplary health app (TrackYourTinnitus) [27]. This video can be requested from the corresponding author. Each rater was trained using this video, and five predefined apps were then rated to ensure that each rater was appropriately trained. If the individual rating score was different from our standard rating score by at least 2 points, the difference was discussed until agreement. All raters had at least a bachelor's degree in psychology to ensure a necessary minimum psychodiagnostic competence standard.

German Version of the Mobile Application Rating Scale

We added the following items in the app description section for the MARS-G: theoretical background (cognitive-behavioral, therapy, systemic therapy, etc), methods (eye movement desensitization and reprocessing, tracking, feedback, etc), category in the app store (lifestyle, medicine, etc), embedding into routine care (communication with therapist, etc), type of use (prevention, treatment, rehabilitation, etc), guidance

(stand-alone, blended care, etc), certification (medical device law, etc), and data safety (log in, informed consent, etc). The four sections of the original MARS were expanded with an additional section focusing on the therapeutic gain associated with the app. The derived items were as follows: gain for the patient; gain for the therapist; risks and adverse effects; and ease of implementation in routine health care.

Analyses

Intraclass Correlation

The included MHAs were rated independently by two trained raters. The intraclass correlation coefficient (ICC) was calculated to assess the extent of agreement between the raters. An ICC of <0.50 indicated poor correlation, $0.51-0.75$ indicated moderate correlation, $0.76-0.89$ indicated good correlation, and >0.90 indicated excellent correlation [28]. According to the findings of previous studies, an ICC >0.75 was considered to indicate sufficient correlation [8,29,30].

Internal Consistency

Internal consistency of the MARS-G and its subscales was assessed as a measure of scale reliability, similar to the original MARS [16]. Omega was used instead of the widely adopted Cronbach alpha to assess reliability, as it provides a more unbiased estimation of reliability [31-33]. For estimations, the procedure introduced by Zhang and Yuan [34] was used to obtain robust coefficients and bootstrapped bias-corrected confidence intervals. Reliability of $\omega <0.50$ indicated unacceptable internal consistency, $0.51-0.59$ indicated poor consistency, $0.60-0.69$ indicated questionable consistency, $0.70-0.79$ indicated acceptable consistency, $0.80-0.89$ indicated good consistency, and >0.90 indicated excellent consistency [35].

Validity

We assessed correlations between corresponding subscales of the MARS and MARS-G, as well as the overall correlation between the MARS total score and MARS-G total score. A r value >0.8 was a priori considered by the author group as an indicator of a strong and sufficient association between the MARS and MARS-G. Additionally, mean comparisons were performed between the dimensions of the MARS and MARS-G, using two-sided t tests. For all comparisons, a P value $<.05$ was considered significant.

Mokken Scale Analysis

Mokken scale analysis (MSA) is a scaling approach closely related to nonparametric item response theory [36]. The preconditions to use MSA are monotonicity and nonintersection. The key parameter in the MSA is Loevinger H_i . H_i is the scaling parameter for item i , and the overall scalability of all items clustering onto scale k is H_k . H_i indicates the strength of the relationship between a latent variable (app quality) and item i . A high scalability score indicates a high probability that an increase in item i is accompanied by an increase in the latent variable. A scale is considered weak if H is <0.4 , moderate if H is ≥ 0.4 but <0.49 , and strong if H is >0.5 [36]. This approach has been described in detail previously [36-39]. For both the MARS and MARS-G, the MSA was conducted to assess the

scalability of the mean scores. As recommended by van der Ark [36], the reliability of the scales was additionally assessed using the Molenaar-Sijtsma method (MS) [40,41], lambda-2 [42], and latent class reliability coefficient (LCRC) [43]. The MSA has been described previously [36].

Analysis Software

R software (R Foundation for Statistical Computing, Vienna, Austria) [44] was used for all analyses, except intraclass correlation. The MSA was conducted using the R package *mokken* [36,38]. Correlations and internal consistency were calculated using the *psych* (version 1.8.12) [45] and *coefficientalpha* packages (version 0.5) [34]. The *coefficientalpha* package includes the calculation of omega with missing and nonnormal data. The ICC was calculated using IBM SPSS 24 (IBM Corp, Armonk, New York) [46].

Results

Descriptive Data and Mean Comparisons

The ICCs for the MARS and MARS-G were high (ICC_{MARS} : 0.84, 95% CI 0.82-0.85; $ICC_{\text{MARS-G}}$: 0.83, 95% CI 0.82-0.85). The mean and standard deviation scores of the items in the MARS-G are presented in Table 1. The mean and standard deviation scores of the items in the MARS are reported elsewhere (E M Messner et al, unpublished data, 2020). The mean scores of the dimensions *engagement* ($t_{206}=0.12$; $P=.91$), *functionality* ($t_{205}=0.39$; $P=.70$), *esthetics* ($t_{206}=-0.012$; $P=.99$), and *information quality* ($t_{204}=0.45$; $P=.66$) and the overall rating ($t_{206}=0.27$; $P=.80$) were equivalent between the MARS and MARS-G.

Table 1. Summary of item and scale scores for the German version of the Mobile App Rating Scale.

Dimension	Score, mean (SD)
Engagement	2.52 (0.70)
Item 01	2.64 (0.93)
Item 02	2.79 (0.90)
Item 03	2.19 (1.00)
Item 04	1.86 (0.79)
Item 05	3.15 (0.72)
Functionality	4.12 (0.69)
Item 06	4.13 (0.82)
Item 07	4.24 (0.77)
Item 08	4.09 (0.74)
Item 09	4.03 (0.78)
Esthetics	3.21 (0.94)
Item 10	3.40 (0.93)
Item 11	3.20 (1.09)
Item 12	3.04 (0.99)
Information quality	2.75 (0.60)
Item 13	3.60 (0.76)
Item 14	2.63 (0.68)
Item 15	2.67 (0.76)
Item 16	2.61 (0.88)
Item 17	3.66 (0.68)
Item 18	1.87 (0.89)
Item 19	3.00 (N/A ^a)
Overall mean	3.11 (0.58)

^aThis item on information quality could be rated for only 1 app, for the rest it was rated *not applicable*.

Internal Consistency

The internal consistency for the MARS dimension *engagement* was good ($\omega=0.84$, 95% CI 0.77-0.88). The internal consistencies for *functionality* ($\omega=0.90$, 95% CI 0.85-0.94) and

esthetics ($\omega=0.91$, 95% CI 0.92-0.96) were excellent. The internal consistency for *information quality* was acceptable ($\omega=0.74$, 95% CI 0.14-0.99; $\alpha=.75$, 95% CI 0.67-0.83). The internal consistency of the overall MARS score was good ($\omega=0.81$, 95% CI 0.74-0.86).

The internal consistencies of the MARS-G dimensions were almost identical to those of the original MARS (engagement: $\omega=0.85$, 95% CI 0.78-0.89; functionality: $\omega=0.91$, 95% CI 0.87-0.94; esthetics: $\omega=0.93$, 95% CI 0.90-0.95; information quality: $\omega=0.72$, 95% CI 0.33-0.81). The internal consistency of the overall score was good ($\omega=0.82$, 95% CI 0.76-0.86).

Validity

The correlation coefficients between corresponding dimensions of the MARS and MARS-G ranged from 0.93 to 0.98, and P values were adjusted for multiple testing according to the Holmes method [47] (Table 2). Correlations between the respective items are presented in Multimedia Appendix 1. There were no associations between user ratings and quality ratings (Table 1).

Table 2. Validity of the German version of the Mobile App Rating Scale (r and P value).

Dimension	Engagement _{GER} ^a	Functionality _{GER}	Esthetics _{GER}	Information quality _{GER}	Star rating
Engagement _{ENG} ^b	0.97 (<.001)	0.49 (<.001)	0.73 (<.001)	0.52 (.001)	-0.03 (.99)
Functionality _{ENG}	0.45 (<.001)	0.98 (<.001)	0.43 (<.001)	0.36 (.002)	0.06 (.99)
Esthetics _{ENG}	0.69 (<.001)	0.41 (<.001)	0.97 (<.001)	0.41 (.001)	0.12 (.99)
Information quality _{ENG}	0.55 (<.001)	0.34 (.004)	0.47 (<.001)	0.93 (.001)	0.25 (.19)
Star rating	-0.03 (>.99)	0.07 (>.99)	0.12 (>.99)	0.26 (.19)	— ^c

^aGerman version.

^bEnglish version.

^cNot applicable.

Mokken Scale Analysis

The MSA of the MARS revealed strong scalability ($H=0.50$; SE 0.062). There were no violations of monotonicity and nonintersection. The internal consistency of this scale was

acceptable ($MS=0.74$; $\lambda_2=0.73$; $LCRC=0.72$). The MSA of the MARS-G revealed good scalability ($H=0.48$; SE 0.060). The internal consistency of this scale was acceptable ($MS=0.74$; $\lambda_2=0.72$; $LCRC=0.74$). The scalability results of the MARS and MARS-G are presented in Table 3.

Table 3. Summary of the Hk coefficient (overall scalability of all items in the scale) for the Mobile App Rating Scale (MARS) and the German version of the Mobile App Rating Scale (MARS-G).

Dimension	MARS	MARS-G
Engagement	0.59	0.57
Functionality	0.43	0.41
Esthetics	0.51	0.51
Information quality	0.45	0.41
Total scale	0.50	0.48

Discussion

Principal Findings

This study developed and evaluated the MARS-G for MHAs. The results showed that the MARS-G is a reliable and valid tool for experts to assess the quality of MHAs. The validity and reliability of the MARS-G were comparable to those of the original MARS. With regard to the reliability of the dimension *information quality*, the confidence interval of omega was overestimated owing to planned missingness. The planned missingness originated from the response option *not applicable*, which allows raters to skip an item if the app does not have any health information (eg, diary apps and brain games). There were no differences in reliability between the MARS-G and original MARS.

The MSA revealed that the use of the MARS-G total score is appropriate. Furthermore, there was good correspondence

between the MARS-G and original MARS, indicating good validity. Our results are consistent with the findings of a study that introduced and tested an Italian version of the MARS [25].

The MARS-G has been presented in Multimedia Appendix 2 and can be obtained from the authors on request. It can be used freely for research and noncommercial MHA-evaluation projects. To reach satisfactory interrater reliability, completion of an online training exercise provided by the corresponding author is highly recommended. Furthermore, a training dataset of five apps can be obtained from the corresponding author on request. The MARS-G ratings should be revised until an appropriate level (ie, ICC >0.75) of interrater reliability is achieved.

To assist in MHA selection, standardized high-quality ratings of MHA are needed in German-speaking countries. Overall, a publicly available database presenting reliable, valid, and standardized expert ratings, like MARS-G ratings, could

contribute to informed health care decisions on which app to use for a specific disease or purpose. The mobile health app database [48] is one example of such a tool that assists users and health care providers in selecting appropriate apps for different health-related purposes.

Limitations

This study has several limitations. First, convergent validity was only evaluated by comparing the MARS and MARS-G. Comparisons with other app rating scales, such as ENLIGHT [15] and the American Psychological Association app evaluation model [12], are necessary in future studies. Second, the focus on anxiety apps limits generalization. Further studies are needed to confirm that these findings can be generalized to other mobile health domains. Such studies would require expert raters who are familiar with the specific domain. Finally, a confirmatory factor analysis of the MARS and MARS-G should be conducted in future studies with larger samples to ensure that the predefined subscales of the MARS and MARS-G can be confirmed.

Future Research

This translation study of the MARS led to the discovery of several research gaps. Future studies should focus on the improvement of app quality assessment and therefore the augmentation of safe MHA use on a broad scale. A challenge in this research is that the sequence in which apps are presented in the app store is incomprehensible and differs depending on which account is used for the search. In future studies, a web

crawler could be used to search European app stores with keywords in order to build an unbiased database of available MHAs. Such a database already exists in China, and it contains all MHAs available in the United States, China, Japan, Brazil, and Russia [49].

Future studies should also shed light on the correlation between real-life user behavior and MARS or MARS-G ratings. As the MARS and MARS-G capture app quality, they could help predict the ability of users to download and use digital resources. Such research has already been conducted for ENLIGHT and real-life user engagement [50]. The efficacy of MHAs is strongly related to user adherence [50-52]; thus, high-quality apps might need to include adherence facilitation strategies to reach their potential.

Moreover, patient involvement should be taken into account. The user version of the MARS (uMARS) [53] should be translated and tested for reliability and validity as well, so that expert ratings of the MARS-G can be complemented with user ratings of the uMARS-G in German-speaking countries. In addition, there is a need for additional studies in the future to investigate the MARS-G and uMARS-G for apps related to specific health problems.

In conclusion, the MARS-G could be used by various stakeholders, such as public health authorities, patient organizations, researchers, health care providers (eg, physicians and psychotherapists), and interested third parties, to assess MHA quality. Furthermore, app developers could use the MARS-G as a tool to improve the quality of their apps.

Acknowledgments

The authors thank Linda Maria Zisch for her help in the translation process.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Item correlation matrix of MARS and MARS-G.
[DOCX File, 25 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Mobile Application Rating Scale-German.
[PDF File (Adobe PDF File), 563 KB-Multimedia Appendix 2]

References

1. UM London, eMarketer. Statista - The Statistics Portal. 2014 Aug. Smartphone user penetration as percentage of total population in Western Europe from 2011 to 2018 URL: <https://www.statista.com/statistics/203722/smartphone-penetration-per-capita-in-western-europe-since-2000/> [accessed 2019-12-05]
2. ForwardAdGroup. Statista - Das Statistik-Portal. 2015. Wie viele Apps haben Sie auf Ihrem Smartphone installiert? URL: <https://de.statista.com/statistik/daten/studie/162374/umfrage/durchschnittliche-anzahl-von-apps-auf-dem-handy-in-deutschland/> [accessed 2019-12-05]
3. Thranberend T, Knöppler K, Neisecke T. Gesundheits-Apps: Bedeutender Hebel für Patient Empowerment - Potenziale jedoch bislang kaum genutzt. Spotlight Gesundh 2016;2:1-8.
4. Deutschland.de. Deutschland.de. 2018. We speak German URL: <https://www.deutschland.de/en/topic/culture/the-german-language-surprising-facts-and-figures> [accessed 2019-04-24]

5. Contributors Wikipedia. Wikipedia, The Free Encyclopedia. 2019. List of languages by number of native speakers URL: https://en.wikipedia.org/wiki/List_of_languages_by_number_of_native_speakers [accessed 2019-04-24]
6. Ellis E, Gogolin I, Clyne M. The Janus Face of Monolingualism: A Comparison of German and Australian Language Education Policies. *Curr Issues Lang Plan* 2010;11(4):439-460. [doi: [10.1080/14664208.2010.550544](https://doi.org/10.1080/14664208.2010.550544)] [Medline: [26281194](https://pubmed.ncbi.nlm.nih.gov/26281194/)]
7. Heron KE, Smyth JM. Ecological momentary interventions: Incorporating mobile technology into psychosocial and health behaviour treatments. *Br J Health Psychol* 2010 Feb;15(1):1-39 [FREE Full text] [doi: [10.1348/135910709X466063](https://doi.org/10.1348/135910709X466063)] [Medline: [19646331](https://pubmed.ncbi.nlm.nih.gov/19646331/)]
8. Terhorst Y, Rathner EM, Baumeister H, Sander L. «Hilfe aus dem App-Store?»: Eine systematische Übersichtsarbeit und Evaluation von Apps zur Anwendung bei Depressionen. *Verhaltenstherapie* 2018 May 8;28(2):101-112. [doi: [10.1159/000481692](https://doi.org/10.1159/000481692)]
9. Ebert DD, Van Daele T, Nordgreen T, Karekla M, Compare A, Zarbo C, et al. Internet and mobile-based psychological interventions: Applications, efficacy and potential for improving mental health. *Eur Psychol* 2018 Jul;23(2):167-187. [doi: [10.1027/1016-9040/a000346](https://doi.org/10.1027/1016-9040/a000346)]
10. Boulos MNK, Brewer AC, Karimkhani C, Buller DB, Dellavalle RP. Mobile medical and health apps: state of the art, concerns, regulatory control and certification. *Online J Public Health Inform* 2014 Feb;5(3):229 [FREE Full text] [doi: [10.5210/ojphi.v5i3.4814](https://doi.org/10.5210/ojphi.v5i3.4814)] [Medline: [24683442](https://pubmed.ncbi.nlm.nih.gov/24683442/)]
11. Albrecht UV. Kapitel 8. Gesundheits-Apps und Risiken. In: Albrecht UV, editor. Chancen und Risiken von Gesundheits-Apps (CHARISMHA). Hannover: Medizinische Hochschule Hannover; 2016:176-192.
12. American Psychiatric Association. American Psychiatric Association. 2017. App evaluation model URL: <https://www.psychiatry.org/psychiatrists/practice/mental-health-apps/app-evaluation-model> [accessed 2019-12-05]
13. Boudreaux ED, Waring ME, Hayes RB, Sadasivam RS, Mullen S, Pagoto S. Evaluating and selecting mobile health apps: Strategies for healthcare providers and healthcare organizations. *Transl Behav Med* 2014 Dec;4:363-371 [FREE Full text] [doi: [10.1007/s13142-014-0293-9](https://doi.org/10.1007/s13142-014-0293-9)] [Medline: [25584085](https://pubmed.ncbi.nlm.nih.gov/25584085/)]
14. Nouri R, Kalhori S, Ghazisaeedi M, Marchand G, Yasini M. Criteria for assessing the quality of mHealth apps: a systematic review. *J Am Med Informatics Assoc* 2018;1-10 [FREE Full text] [doi: [10.1093/jamia/ocy050](https://doi.org/10.1093/jamia/ocy050)]
15. Baumel A, Faber K, Mathur N, Kane JM, Muench F. Enlight: A Comprehensive Quality and Therapeutic Potential Evaluation Tool for Mobile and Web-Based eHealth Interventions. *J Med Internet Res* 2017 Mar 21;19(3):e82 [FREE Full text] [doi: [10.2196/jmir.7270](https://doi.org/10.2196/jmir.7270)] [Medline: [28325712](https://pubmed.ncbi.nlm.nih.gov/28325712/)]
16. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. *JMIR mHealth uHealth* 2015 Mar;3(1):e27 [FREE Full text] [doi: [10.2196/mhealth.3422](https://doi.org/10.2196/mhealth.3422)] [Medline: [25760773](https://pubmed.ncbi.nlm.nih.gov/25760773/)]
17. Bardus M, van Beurden SB, Smith JR, Abraham C. A review and content analysis of engagement, functionality, aesthetics, information quality, and change techniques in the most popular commercial apps for weight management. *Int J Behav Nutr Phys Act* 2016 Mar 10;13:35 [FREE Full text] [doi: [10.1186/s12966-016-0359-9](https://doi.org/10.1186/s12966-016-0359-9)] [Medline: [26964880](https://pubmed.ncbi.nlm.nih.gov/26964880/)]
18. Grainger R, Townsley H, White B, Langlotz T, Taylor WJ. Apps for People With Rheumatoid Arthritis to Monitor Their Disease Activity: A Review of Apps for Best Practice and Quality. *JMIR Mhealth Uhealth* 2017 Feb 21;5(2):e7 [FREE Full text] [doi: [10.2196/mhealth.6956](https://doi.org/10.2196/mhealth.6956)] [Medline: [28223263](https://pubmed.ncbi.nlm.nih.gov/28223263/)]
19. Knitza J, Tascilar K, Messner EM, Meyer M, Vossen D, Pulla A, et al. German Mobile Apps in Rheumatology: Review and Analysis Using the Mobile Application Rating Scale (MARS). *JMIR Mhealth Uhealth* 2019 Aug 05;7(8):e14991 [FREE Full text] [doi: [10.2196/14991](https://doi.org/10.2196/14991)] [Medline: [31381501](https://pubmed.ncbi.nlm.nih.gov/31381501/)]
20. Machado GC, Pinheiro MB, Lee H, Ahmed OH, Hendrick P, Williams C, et al. Smartphone apps for the self-management of low back pain: A systematic review. *Best Practice & Research Clinical Rheumatology* 2016 Dec;30(6):1098-1109. [doi: [10.1016/J.BERH.2017.04.002](https://doi.org/10.1016/J.BERH.2017.04.002)]
21. Mani M, Kavanagh DJ, Hides L, Stoyanov SR. Review and Evaluation of Mindfulness-Based iPhone Apps. *JMIR Mhealth Uhealth* 2015 Aug 19;3(3):e82 [FREE Full text] [doi: [10.2196/mhealth.4328](https://doi.org/10.2196/mhealth.4328)] [Medline: [26290327](https://pubmed.ncbi.nlm.nih.gov/26290327/)]
22. Masterson Creber RM, Maurer MS, Reading M, Hiraldo G, Hickey KT, Iribarren S. Review and Analysis of Existing Mobile Phone Apps to Support Heart Failure Symptom Monitoring and Self-Care Management Using the Mobile Application Rating Scale (MARS). *JMIR Mhealth Uhealth* 2016 Jun 14;4(2):e74 [FREE Full text] [doi: [10.2196/mhealth.5882](https://doi.org/10.2196/mhealth.5882)] [Medline: [27302310](https://pubmed.ncbi.nlm.nih.gov/27302310/)]
23. Salazar A, de Sola H, Failde I, Moral-Munoz JA. Measuring the Quality of Mobile Apps for the Management of Pain: Systematic Search and Evaluation Using the Mobile App Rating Scale. *JMIR Mhealth Uhealth* 2018 Oct 25;6(10):e10718 [FREE Full text] [doi: [10.2196/10718](https://doi.org/10.2196/10718)] [Medline: [30361196](https://pubmed.ncbi.nlm.nih.gov/30361196/)]
24. Thornton L, Quinn C, Birrell L, Guillaumier A, Shaw B, Forbes E, et al. Free smoking cessation mobile apps available in Australia: a quality review and content analysis. *Aust N Z J Public Health* 2017 Dec;41(6):625-630. [doi: [10.1111/1753-6405.12688](https://doi.org/10.1111/1753-6405.12688)] [Medline: [28749591](https://pubmed.ncbi.nlm.nih.gov/28749591/)]
25. Domnich A, Arata L, Amicizia D, Signori A, Patrick B, Stoyanov S, et al. Development and validation of the Italian version of the Mobile Application Rating Scale and its generalisability to apps targeting primary prevention. *BMC Med Inform Decis Mak* 2016 Jul 7;16(83):1-10. [doi: [10.1186/s12911-016-0323-2](https://doi.org/10.1186/s12911-016-0323-2)]

26. Martin Payo R, Fernandez Álvarez MM, Blanco Díaz M, Cuesta Izquierdo M, Stoyanov SR, Llana Suárez E. Spanish adaptation and validation of the Mobile Application Rating Scale questionnaire. *Int J Med Inform* 2019 Sep;129:95-99. [doi: [10.1016/j.ijmedinf.2019.06.005](https://doi.org/10.1016/j.ijmedinf.2019.06.005)]
27. Pryss R, Probst T, Schlee W, Schobel J, Langguth B, Neff P, et al. Prospective crowdsensing versus retrospective ratings of tinnitus variability and tinnitus-stress associations based on the TrackYourTinnitus mobile platform. *Int J Data Sci Anal* 2019;327-338. [doi: [10.1007/s41060-018-0111-4](https://doi.org/10.1007/s41060-018-0111-4)]
28. Portney LG, Watkins MP. *Foundations of clinical research: applications to practice*. Upper Saddle River, NJ: Pearson/Prentice Hall; 2009.
29. Lin J, Sander L, Paganini S, Schlicker S, Ebert D, Berking M, et al. Effectiveness and cost-effectiveness of a guided internet- and mobile-based depression intervention for individuals with chronic back pain: Protocol of a multi-centre randomised controlled trial. *BMJ Open* 2017 Dec 28;7:e015226. [doi: [10.1136/bmjopen-2016-015226](https://doi.org/10.1136/bmjopen-2016-015226)]
30. Sander L, Paganini S, Lin J, Schlicker S, Ebert DD, Buntrock C, et al. Effectiveness and cost-effectiveness of a guided Internet- and mobile-based intervention for the indicated prevention of major depression in patients with chronic back pain—study protocol of the PROD-BP multicenter pragmatic RCT. *BMC Psychiatry* 2017 Jan 21;17(36). [doi: [10.1186/s12888-017-1193-6](https://doi.org/10.1186/s12888-017-1193-6)]
31. Dunn TJ, Baguley T, Brunsden V. From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *Br J Psychol* 2014 Aug;105(3):399-412. [doi: [10.1111/bjop.12046](https://doi.org/10.1111/bjop.12046)] [Medline: [24844115](https://pubmed.ncbi.nlm.nih.gov/24844115/)]
32. Revelle W, Zinbarg RE. Coefficients Alpha, Beta, Omega, and the GLB: Comments on Sijsma. *Psychometrika* 2009;74(1):145-154. [doi: [10.1007/s11336-008-9102-z](https://doi.org/10.1007/s11336-008-9102-z)]
33. McNeish D. Thanks Coefficient Alpha, We'll Take It From Here. *Psychol Methods* 2018 Sep;23(3):412-433. [doi: [10.1037/met0000144](https://doi.org/10.1037/met0000144)]
34. Zhang Z, Yuan KH. Robust Coefficients Alpha and Omega and Confidence Intervals With Outlying Observations and Missing Data: Methods and Software. *Educ Psychol Meas* 2016 Jun;76(3):387-411 [FREE Full text] [doi: [10.1177/0013164415594658](https://doi.org/10.1177/0013164415594658)] [Medline: [29795870](https://pubmed.ncbi.nlm.nih.gov/29795870/)]
35. George D, Mallery P. *SPSS For Windows Step By Step: A Simple Guide And Reference, 11.0 Update*. Boston: Allyn & Bacon; 2003.
36. van der Ark LA. Mokken Scale Analysis in R. *J Stat Softw* 2007 Nov 8;20(11):1-19. [doi: [10.1007/s11336-007-9034-z](https://doi.org/10.1007/s11336-007-9034-z)]
37. Mokken RJ. *A theory and procedure of scale analysis: With applications in political research*. New York: De Gruyter Mouton; 1971.
38. van der Ark LA. New Developments in Mokken Scale Analysis in R. *J Stat Softw* 2012;48(5):1-27. [doi: [10.18637/jss.v048.i05](https://doi.org/10.18637/jss.v048.i05)]
39. Sijsma K, van der Ark LA. A tutorial on how to do a Mokken scale analysis on your test and questionnaire data. *Br J Math Stat Psychol* 2017;70(1):137-158. [doi: [10.1111/bmsp.12078](https://doi.org/10.1111/bmsp.12078)]
40. Molenaar I, Sijsma K. Mokken's Approach to Reliability Estimation Extended to Multicategory Items. *Kwant Methoden* 1988;9(28):115-126.
41. Sijsma K, Molenaar IW. Reliability of test scores in nonparametric item response theory. *Psychometrika* 1987 Mar;52(1):79-97. [doi: [10.1007/bf02293957](https://doi.org/10.1007/bf02293957)]
42. Guttman L. A basis for analyzing test-retest reliability. *Psychometrika* 1945 Dec;10(4):255-282. [doi: [10.1007/bf02288892](https://doi.org/10.1007/bf02288892)]
43. van der Ark LA, van der Palm DW, Sijsma K. A Latent Class Approach to Estimating Test-Score Reliability. *Appl Psychol Meas* 2011 Mar 09;35(5):380-392. [doi: [10.1177/0146621610392911](https://doi.org/10.1177/0146621610392911)]
44. R Core Team. *R: A Language and Environment for Statistical Computing*. R Found Stat Comput Vienna, Austria 2017 [FREE Full text]
45. Revelle W. *psych: Procedures for Psychological, Psychometric, and Personality Research [Computer Software]*. 2017. URL: <https://personality-project.org/r/psych> [accessed 2019-11-22]
46. IBM Corporation. *IBM SPSS Advanced Statistics 24 [Software]*. 2016. URL: http://www-01.ibm.com/support/docview.wss?uid=swg27047033#ja%5Cnftp://public.dhe.ibm.com/software/analytics/spss/documentation/statistics/24.0/ja/client/Manuals/IBM_SPSS_Advanced [accessed 2020-01-16]
47. Holm S. A simple sequentially rejective multiple test procedure. *Scand J Stat* 1979;6:65-70 [FREE Full text]
48. MHAD Core Team. *Mobile Health App Database*. 2019. URL: <http://www.mhad.science/> [accessed 2019-09-11]
49. Xu W, Liu Y. mHealthApps: A Repository and Database of Mobile Health Apps. *JMIR mHealth uHealth* 2015 Mar 18;3(1):e28 [FREE Full text] [doi: [10.2196/mhealth.4026](https://doi.org/10.2196/mhealth.4026)] [Medline: [25786060](https://pubmed.ncbi.nlm.nih.gov/25786060/)]
50. Baumeister A, Yom-Tov E. Predicting user adherence to behavioral eHealth interventions in the real world: examining which aspects of intervention design matter most. *Transl Behav Med* 2018 Sep 08;8(5):793-798. [doi: [10.1093/tbm/ibx037](https://doi.org/10.1093/tbm/ibx037)] [Medline: [29471424](https://pubmed.ncbi.nlm.nih.gov/29471424/)]
51. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression: Systematic review. *J Med Internet Res* 2009 Apr;11(2):e13 [FREE Full text] [doi: [10.2196/jmir.1194](https://doi.org/10.2196/jmir.1194)] [Medline: [19403466](https://pubmed.ncbi.nlm.nih.gov/19403466/)]
52. Van Ballegooijen W, Cuijpers P, Van Straten A, Karyotaki E, Andersson G, Smit JH, et al. Adherence to internet-based and face-to-face cognitive behavioural therapy for depression: A meta-analysis. *PLoS ONE* 2014 Jul;9(7):e100674 [FREE Full text] [doi: [10.1371/journal.pone.0100674](https://doi.org/10.1371/journal.pone.0100674)] [Medline: [25029507](https://pubmed.ncbi.nlm.nih.gov/25029507/)]

53. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and Validation of the User Version of the Mobile Application Rating Scale (uMARS). JMIR mHealth uHealth 2016 Jun 10;4(2):e72 [FREE Full text] [doi: [10.2196/mhealth.5849](https://doi.org/10.2196/mhealth.5849)] [Medline: [27287964](https://pubmed.ncbi.nlm.nih.gov/27287964/)]

Abbreviations

ENLIGHT: Evaluation Tool for Mobile and Web-Based eHealth Interventions

ICC: intraclass correlation coefficient

LCRC: latent class reliability coefficient

MARS: Mobile App Rating Scale

MARS-G: German version of the Mobile App Rating Scale

MHA: mobile health app

MS: Molenaar-Sijtsma method

MSA: Mokken scale analysis

uMARS: user version of the Mobile App Rating Scale

Edited by G Eysenbach; submitted 24.04.19; peer-reviewed by C Aljoscha, M Bardus, E de Krijger, R Bipeta; comments to author 05.06.19; revised version received 29.07.19; accepted 24.09.19; published 27.03.20

Please cite as:

Messner EM, Terhorst Y, Barke A, Baumeister H, Stoyanov S, Hides L, Kavanagh D, Pryss R, Sander L, Probst T

The German Version of the Mobile App Rating Scale (MARS-G): Development and Validation Study

JMIR Mhealth Uhealth 2020;8(3):e14479

URL: <http://mhealth.jmir.org/2020/3/e14479/>

doi: [10.2196/14479](https://doi.org/10.2196/14479)

PMID:

©Eva-Maria Messner, Yannik Terhorst, Antonia Barke, Harald Baumeister, Stoyan Stoyanov, Leanne Hides, David Kavanagh, Rüdiger Pryss, Lasse Sander, Thomas Probst. Originally published in JMIR mHealth and uHealth (<http://mhealth.jmir.org>), 27.03.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on <http://mhealth.jmir.org/>, as well as this copyright and license information must be included.